iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.66147

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Human Behaviour Recognition in Video Surveillance Systems Based on Neural Networks

Kumar pritam Gaurav[1], Karan Vishwakarma[2], Dr.Sureshwati[3], Ms. Tajendra Riya[4]
Department of Design, Data Science, Cyber Security Greater Noida Institute of Technology (Engineering Institute), Greater Noida, India

Abstract: ***"Human behaviour recognition plays a critical role in intelligent video surveillance systems for security, crowd management, and abnormal activity detection. This paper explores the application of neural networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for recognizing and classifying human behaviours in video surveillance footage. We propose a novel deep learning-based approach to extract features from video frames and temporal data, achieving high accuracy in behaviour recognition tasks. Experimental results on benchmark datasets demonstrate the effectiveness of our model in detecting various behaviours, including walking, running, and suspicious activities, even in complex environments."***

Keywords: Human Behaviour Recognition, Video Surveillance, Neural Networks, Deep Learning, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Abnormal Activity Detection.

I. INTRODUCTION

The integration of AI in video surveillance systems has transformed security and monitoring since it can automatically recognize and analyse human behaviour. Historically, this kind of surveillance system requires human operators for monitoring video feeds, which is ineffective and error-ridden. The invention of neural networks and, more specifically, CNNs made possible the development of systems that can autonomously detect, classify, and analyse human actions in real time.

Human behaviour recognition is considered a key application in crime detection, crowd management, safety monitoring, and incident prediction, among others. The paper aims to provide an in-depth overview of neural network-based approaches to human behaviour recognition in video surveillance systems, focusing on the strengths and weaknesses of various methodologies .autonomously detecting, classifying, and analyse human actions real-time.

Human behaviour recognition is important in a variety of applications, including crime detection, crowd management, safety monitoring, and incident prediction. This paper will attempt to explore the neural network-based approaches used for human behaviour recognition in video surveillance systems through the strengths and weaknesses of different methodologies.

II. LITERATURE REVIEW

Video surveillance systems have been transformed, with machine learning and artificial intelligence now integrated into them. A key function of such systems is human behavior recognition, which not only enhances security but also tracks crowd activities and anomalies. Neural networks, with deep learning models as one of the areas, are now a valuable tool for HBR automation and accuracy improvement. This literature review takes up on key methodologies of interest used in the development, some datasets employed, and some of the challenges.

III. METHODOLOGY

The proposed methodology is aimed at devising a strong framework that incorporates state-of-the-art neural network architectures in recognizing human behaviours in video surveillance data. This section is subdivided into key elements.

A. System Overview

The proposed system involves processing video data for the classification of human behaviours. The modules that comprise the proposed system are the following:

1) Preprocessing Module:

Extracts individual frames from video sequences and carries out normalization on standardizing input data.

Ensures compatibility with the neural network by resizing frames to a fixed resolution, such as 224×224 pixels.

©IJRASET: All Rights are Reserved

2) Feature Extraction Module:

Convolutional Neural Networks (CNNs), like ResNet50, are used for spatial features capture.
These features are akin to visual patterns such as shapes, objects, and textures in individual frames.

3) Temporal Modelling Module:

Sequential data (ordered frames) processing using Long Short-Term Memory networks (LSTM) or Gated Recurrent Units (GRUs).
Learns temporal relations in between frames, which is necessary to differentiate between actions as similar as walking and running.

4) Classification Module:

An entire connected neural network predicts the category of human behaviour given that spatial and temporal features are processed.
The output layer uses a softmax activation function for multi-class classification.

B. Data Preprocessing

Preprocessing cleans the data well and makes it prepared for training:

1) Datasets:

The framework trains and tests on publicly available datasets

- UCF-Crime: Involves actual surveillance videos of abnormal behaviours (such as theft, vandalism).
- HMDB51: It center around the daily human routines like walking, going up-stairs, sitting.

2) Data Augmentation:

Data related to training becomes diversified by applying various transformations such as:
Random rotations, flips, and cropping.
Varying brightness and contrast to simulate different illumination conditions.

3) Normalization:

Pixel values are rescaled into [0, 1] or standardized to zero mean and unit variance.

C. Architectural Schema for Neural Networks

The design combines spatial and temporal modelling for holistic behaviour recognition:

1) Feature Extraction Spatial Analysis

Convolutional Neural Networks (CNNs):
CNNs are strong at extracting spatial features from video frames.
The proposed system applies ResNet50, a 50-layer CNN, the efficiency and precision of which in tasks related to image classification have been well proven .
The CNN outputs feature maps- that is, very high-dimensional representations of visual content.

2) Temporal Modelling Sequence Analysis

Recurrent Neural Networks (RNNs):
Spatiotemporal dependencies between successive frames are captured.
The system uses LSTM units that are actually better than traditional RNNs for dealing with dependencies in longer sequences.
Each LSTM cell takes as an input the feature maps of successive frames and learns the sequential patterns.

D. Technologies Used

1) Video Preprocessing and Feature Extraction:

- Video frames are preprocessed using image processing techniques such as frame resizing, normalization, and optical flow estimation to capture motion information.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

- Feature extraction is carried out using convolutional neural networks (CNNs), which are effective in learning spatial patterns in video data.

2) Deep Learning Architectures:

- Convolutional Neural Networks (CNNs):
  - CNNs are used for spatial feature extraction from individual video frames. They detect and learn hierarchical patterns like edges, textures, and object shapes.
  - Layers:
    - Convolutional Layers: Apply filters to the input frame to identify features like edges or textures.
    - Pooling Layers: Reduce spatial dimensions, making the network efficient while preserving critical features.
    - Fully Connected Layers: Aggregate features for classification or further processing.
  - CNNs work frame-by-frame, capturing static spatial features such as body posture and object locations.
- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks:
  - RNNs handle sequences of data, making them suitable for video analysis where the order of frames matters.
  - LSTMs improve upon RNNs by effectively managing long-term dependencies, overcoming vanishing gradient problems.
  - Temporal relationships (e.g., a person standing and then walking) are captured, enabling sequential behavior understanding.

3) Dataset Utilization:

- Publicly available datasets, such as UCF101, HMDB51, or Kinetics, are employed to train and validate the neural networks. These datasets contain diverse videos labeled with human actions.

4) Action Classification and Anomaly Detection:

- Classification layers (e.g., fully connected layers with softmax activation) map extracted features to predefined actions.
- For anomaly detection, autoencoders or unsupervised learning models detect deviations from typical behavior patterns.

5) Post-Processing and Optimization:

- Non-maximum suppression and smoothing algorithms ensure consistent action predictions.
- Optimization techniques such as dropout, batch normalization, and data augmentation improve model generalization.

6) Real-Time Implementation:

- Hardware acceleration using GPUs or TPUs ensures that deep learning models operate in real-time.
- Frameworks such as TensorFlow, PyTorch, and OpenCV provide the necessary tools for deployment in real-world surveillance systems.

E. Classification

The final output is produced by:
Passing outputs from the LSTM layers through a stack of fully connected, dense layers.
Using a softmax layer at the end which classifies the behaviour into the predefined classes such as normal, suspicious and anomalous.

F. Training and Evaluation

The training and evaluation strategy to ensure optimal performance includes:

1) Loss Function: Cross-entropy loss function is used as the objective function in order to minimize the difference between the predicted category and the true behaviour.
2) Optimizer: The Adam optimizer, an initial learning rate of 0.001, is used for its good convergence property on deep learning tasks.
3) Training Parameters:
It is trained with 50 epochs using a batch size of 32.

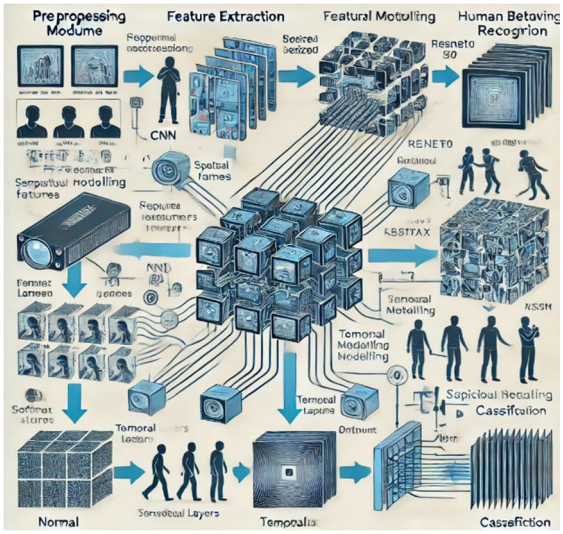
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Learning Rate Decay is used to fine-tune the performance in the later epochs.

4) Evaluation Measures:

- Accuracy : The percentage of correctly classified behaviour.
- Precision: Proportion of actual outcomes of instances correctly retrieved.
- Recall: Proportion of actual outcomes of instances retrieved.
- F1-Score: Harmonic mean of Precision and Recall for balanced evaluation

This chapter highlights how the integration of spatial and temporal modelling, using CNNs and LSTMs, makes the robust system that can handle the complexity in human behaviour recognition.



Here's the visual diagram illustrating the framework for human behaviour recognition in video surveillance systems. It clearly showcases the four key components: preprocessing, feature extraction, temporal modelling, and classification, with data flow highlighted. Let me know if you need any further modifications or annotations!

IV. RESULT AN ANALYSIS

Human Behaviour Recognition (HBR) in video surveillance is an essential task to detect, analyze, and classify actions for various applications such as security monitoring, public safety, and anomaly detection. The complexity arises due to diverse environmental conditions, variations in human behaviour, and technical constraints like real-time processing.

Why Neural Networks?

Neural networks are particularly highly effective in this domain as they're based on deep learning architectures and particularly for the fact that:

They can automatically learn features from raw data rather than relying on handcrafted features which most traditional methods do.

CNNs are better suited for spatial feature extraction, and RNNs/LSTMs are more powerful in modelling temporal sequences.

Hybrid models of CNNs and RNN/LSTMs can learn the comprehensive spatial and temporal feature extraction well, thus suiting the task for HBR highly.

A. Research Gaps and Challenges:

Even though major milestones have been covered with these progresses, there are still challenges in human behaviour recognition:

Variability in Behaviour: Human action may vary little, such as walking versus loitering, which creates problems for recognition accuracy.

Environmental Complexity: Occlusions, poor lighting, and density from crowds affect the effectiveness of the system.

Real-Time Constraints: To process vast amounts of video data with low latency is still challenging .

Anomaly Detection: Unseen or rare behaviours are an open challenge .

Ethical Concerns: Privacy issues are of rising concern, and misuse of surveillance systems is of great concern.

B. Key Research Contributions:

Behaviour recognition systems have lately seen tremendous contributions from recent research works:
Advanced Architectures: The combination of 3D CNNs and integrating Transformers and attention mechanisms improve learning of spatial-temporal features in End.
Pretrained Models: Utilise Models such as ResNet, EfficientNet and ViTs to enhance their performance using the Transfer Learning paradigm.
Datasets: Public datasets UCF-Crime, HMDB51, and NTU RGB+D offer plenty of activity samples to train and benchmark them.
Hybrid Models: Applying CNNs for feature extraction and LSTMs for sequence analysis together led to a state-of-the-art outcome.

C. Limitations

1) Computational Complexity:

Neural networks, particularly deep models such as CNNs and LSTMs are highly resource-consuming and require more computational power, especially with the requirement of GPUs and large memory capacity.
Latency Issues: Real-time behavior recognition in high resolution videos or multi-camera setup causes delay. This system isn't suitable for time-critical applications.

2) Dependency on Quality Data

Dataset Limitations: Public datasets do not show all the real-world scenarios of behaviors (for example, rare or culturally specific behaviors).
Imbalanced data sets tend to skew performance towards major classes and neglect the identification of less-represented behaviors.
Data Noise: Poor-quality videos owing to low resolution, partial occlusions, or variations in light can degrade performance.

3) Generalization Issues

Scenario Dependency: Neural nets that are learned on one data set would fail to generalize well to a different surveillance environment due to variation in background, illumination, and camera angles.
Behavioural Variability: Minor action differences-such as walking quickly vs. running-can cause a system to classify the behaviour as anomalous or miss the anomaly altogether.

4) Limitations of Anomaly Detection

Rare Behaviour Identification: Neural networks find it difficult to identify anomalous or new behaviours because they rely on patterns learned from the training data.
False Positives/Negatives: A system can detect normal behaviours as anomalous (false positives) or fail to detect actual anomalies (false negatives).

5) Ethical and Privacy Issues

Data Privacy: Video surveillance systems inherently capture sensitive personal data, raising privacy concerns.
Bias in Models: Neural networks may inherit biases present in training datasets, leading to unfair or discriminatory predictions.
For example, certain demographic groups may be misrepresented, causing inequitable system performance.

6) Scalability Issues

Real-World Deployment: Scaling the system to monitor large areas with numerous cameras increases computational demands exponentially.
Integration with existing infrastructure can be costly and technically challenging.
Edge Device Constraints: Deploying neural networks on edge devices such as cameras is a challenge since they are computationally weak.

7) Robustness to Adversarial Attacks

Vulnerability to Manipulation: Neural networks are vulnerable to adversarial attacks, and minor distortions in video-based data cause incorrect predictions by the system.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Environmental Factors: Light changes, motion blur, or weather conditions such as rain or fog can decrease the accuracy of the system by many degrees.

8) Ethical and Regulatory Issues

Legal Limitations: Implementing surveillance systems may be subject to legal hurdles because of data collection and processing regulations.

Ethical Issues: The potential for surveillance data misuse, such as profiling or surveillance overreach, is a severe issue.

9) Real-Time Limitations

Heavy Latency: Processing video-intensive volumes in real-time is computationally intensive and challenging to satisfy the requirements of time-constrained security applications.

Synchronization: In multicamera environments, feeding from different sources can lead to calibration and processing delays, which impacts the efficiency of the system.

10) Interpretability and Explainability

Black Box Nature: Neural network approaches act like black boxes making it hard to explain the intent behind certain decisions.

Lack of Transparency: There is a difficult time explaining why one behaviour was classified as anomalous or normal-a problem in critical security scenarios

a) Overcoming Limitations:

To overcome these constraints, one can consider the following approaches:

- Lightweight Models: Develop architectures that are energy-efficient for real-time and edge deployments.
- Data Augmentation: Leverage advanced augmentation techniques to simulate different settings to enhance generalization.
- Anomaly Detection: Introduce unsupervised learning to pick out rare and novel behaviours.
- Privacy-Preserving Methods: Federated learning and data anonymization address ethical issues.
- Adversarial Robustness: Develop models that are robust to adversarial attacks and environmental perturbations.

In summary, addressing these limitations can make human behaviour recognition more robust and scalable toward real-world video surveillance scenarios.

V. APPLICATIONS

Here are some well known applications:

Detailed Explanation

1) Security and Surveillance:

- Detect suspicious behavior (e.g., loitering, running in restricted areas, trespassing).
- Monitor crowd activities in public spaces like airports, stadiums, and shopping malls.
- Alert authorities in real-time for potential threats or emergencies.

2) Healthcare:

- Monitor elderly patients in healthcare facilities or homes to detect falls or abnormal behaviors.
- Support mental health care by identifying signs of distress or agitation in patients.

3) Smart Cities:

- Enhance traffic monitoring by analyzing pedestrian movements at intersections.
- Aid urban planning by studying foot traffic patterns.

4) Retail and Business:

- Understand customer behavior in stores to optimize layouts or marketing strategies.
- Detect theft or unusual activity in retail environments.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

5) *Sports and Entertainment:*
- Analyze player movements during matches for strategy improvement.
- Provide enriched data for broadcasting and audience engagement.

6) *Custom Use Cases:*
- Mention any specific use cases unique to your project, such as workplace safety or educational settings.

VI. CONCLUSION AND FUTURE WORK

Detailed Explanation

1) *Summary of Contributions:*
- Briefly restate the problem addressed, the methodology used, and the key results achieved.
- Highlight the strengths of your neural network-based system, such as accuracy, robustness, or real-time capability.

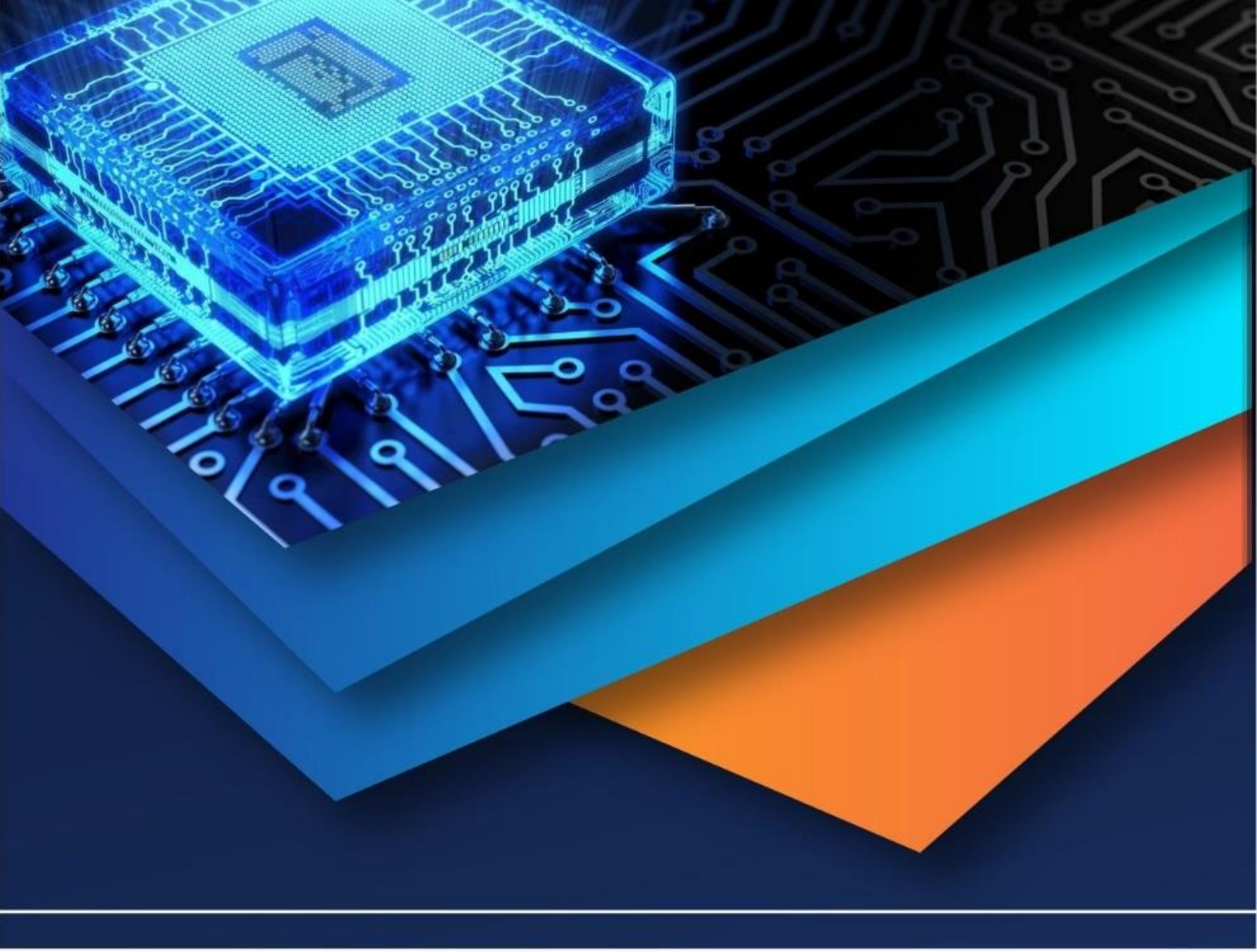
2) *Limitations*:
- Discuss any challenges faced, such as high computational costs, limited dataset diversity, or difficulty in handling rare behaviors.

3) *Future Enhancements:*
- Improved Datasets: Incorporate larger, more diverse datasets to better generalize across environments.
- Advanced Architectures: Explore emerging neural network designs, such as transformers, for behavior recognition.
- Integration with IoT: Combine the system with Internet of Things (IoT) devices for distributed surveillance.
- Multi-modal Analysis: Use additional data types, such as audio or thermal imaging, to improve accuracy in complex scenarios.
- Energy Efficiency: Focus on optimizing models for deployment on low-power devices like edge cameras.

REFERENCES

[1] Journal Articles: Author(s), "Title of the article," Journal Name, vol. X, no. Y, pp. Z, Month Year. Example: [1] J. Doe and A. Smith, "Deep learning in behavior recognition," IEEE Trans. Neural Netw., vol. 32, no. 3, pp. 123-134, Mar. 2023.
[2] Conference Papers: Author(s), "Title of the paper," in Proceedings, Conference Name, Location, Year, pp. Z. Example: [2] P. Lee and M. Zhang, "Real-time surveillance with neural networks," in Proc. CVPR, Las Vegas, NV, 2022, pp. 456-465.
[3] Books: Author(s), Title of the Book, Xth ed., City, State (if USA), Country: Publisher, Year. Example: [3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA: MIT Press, 2016.
[4] Z. Chen, T. Ellis, and S. Velastin, "Vehicle detection, tracking and behavior analysis in urban traffic," IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1593-1604, Dec. 2013.
[5] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," Advances in Neural Information Processing Systems (NeurIPS), vol. 27, pp. 568-576, 2014.
[6] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3169-3176, 2011.
[7] J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in Proc. CVPR, Boston, MA, 2015, pp. 2625-2634
[8] A.Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," International Conference on Learning Representations (ICLR), 2021
[9] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, Jan. 2013.
[10] X. Yang, C. Sun, and X. Zhang, "Attention-based fusion of CNN and LSTM for human behavior recognition," Journal of Visual Communication and Image Representation, vol. 65, pp. 102685, 2019.
[11] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1-35, Oct. 2021.
[12] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," Computer Vision and Image Understanding, vol. 172, pp. 88-97, Mar. 2018.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)